Article

Structural Stability of Intelectin-1

John J. Kozak, Harry B. Gray, and Roberto A. Garza-López

J. Phys. Chem. B, **Just Accepted Manuscript** • DOI: 10.1021/acs.jpcb.6b08691 • Publication Date (Web): 26 Oct 2016

Structural Stability of Intelectin-1

John J. Kozak [a], Harry B. Gray [b] and Roberto A. Garza-López [c,*]

a) DePaul University, 243 South Wabash Ave., Chicago IL 60604-6116.
b) Beckman Institute, California Institute of Technology, Pasadena, CA 91125
c) Department of Chemistry and Seaver Chemistry Laboratory,
Pomona college, Claremont, CA 91711

[*]Corresponding author: rgarza@pomona.edu

Abstract

We study the structural stability of helical and non-helical regions in chain A of human intelectin-1. Using a geometrical model introduced previously, a computational analysis based on the recently reported crystal structure of this protein by Kiessling et al. is carried out to quantify the resiliency of the native state to steric perturbations. Response to these perturbations is characterized by calculating, relative to the native state, the lateral, radial and angular displacements of n-residue segments of the polypeptide chain centered on each residue. By quantifying the stability of the protein through six stages of unfolding, we are able to identify regions in chain A of intelectin-1 which are markedly affected by structural perturbations versus those which are relatively unaffected, the latter suggesting that the native-state geometry of these regions is essentially conserved. Importantly, residues in the vicinity of calcium ions comprise a conserved region, suggesting that Ca ions play a role not only in the coordination of carbohydrate hydroxyl groups, but in preserving the integrity of the structure.

I. Introduction

We have developed a geometrical model to study the role of steric factors in influencing the structural stability and unfolding of proteins.[1-4] Recent work has established that intelectin-1 has properties necessary to function in the immune system's surveillance complex.[5,6] In this study, we examine the response of chain A of intelectin-1 to structural perturbations. In contrast to the azurins [1-3] and cytochromes [4] studied previously, intelectin-1 has (many) more residues and, importantly, three metal (Ca) ions. Kiessling et al. who identify residues involved in calcium coordination and ligand binding, suggest [5,6] that calcium ions could facilitate a kind of "lock-and-key" binding that is often conserved in carbohydrate-binding sites. As the role of coordinated metal ions in stabilizing native polypeptide folds is of great current interest [7-21], we pay particular attention in our geometrical analysis to the regions around the three calciums. We quantify changes in the structure of intelectin-1 in the vicinity of Ca ions as the native state is disrupted.

The geometrical mechanics of our model, described in these references [1-4], will not be reproduced here. Rather, we provide below a summary of our approach, highlighting the main features.

First, we keep intact *nearest-neighbor* repulsive *and* attractive interactions between each residue i and its two nearest-neighbors by introducing a triplet modular unit centered on residue i . The geometry of each triplet (a module of n=3 residues) is determined by the crystallographic evidence. By enforcing the constraint that *nearest-neighbor* interactions defining the native state remain unchanged, the locally-optimized structure of the protein is strictly conserved.

For second nearest-neighbor (and down-range) interactions, we focus entirely on the geometrical constraints which delimit the number of available configurations accessible to a polypeptide chain. We introduce structural perturbations by relaxing incrementally *steric* constraints (*only*) between and among *non*-nearest neighbors. It is in focusing on "excluded volume" effects that our approach is in the same genre as Ramachandran's analysis [22] phi/psi plots introduced in the early sixties . However, it has long been recognized, since van der Waals, that excluded volume effects, purely repulsive interactions, are of primary importance in establishing the structure and properties of a fluid state. For a recent review, see Ref. (23). In following the consequences of relaxing sequentially the geometrical constraints on a globally-optimized structure, the native state of a protein, we quantify one possible pathway down the "folding funnel" to a more disorganized, fluid-like state.

Among the many configurations that can be adopted by a segment of the polypeptide chain as the native state is disrupted , one (uniquely defined) "limiting case" can be identified: a maximally-extended, linear sequence of n residues. Here, as in our previous studies,[1-4] this configuration is taken as representative of one (possible) sterically-perturbed state.

A single triplet modular unit is defined by n=3 residues. Two adjacent triplets, connected in a linear array, involve n=5 residues; three adjacent triplets, connected in a linear array, involve n=7 residues, and so on. In this study, we consider up to (and including) an extended (linear) array of seven adjacent triplets (n=15 residues). In interpreting our data, we shall take advantage of the fact that 3.6 residues comprise one convolution of an α-helix Thus, the n=5,7 linear extensions correspond to the breaking of one H-bond, the n=9 extension corresponds to

breaking two H-bonds, n=11,13 to three H-bonds, and n=15, to four H-bonds. Following changes (such as bond breakage) in the local environment of each residue as the protein unfolds allows us to describe, sequentially, one possible pathway down a "folding funnel" from the native to a fully denatured state.

In our approach, we identify three signatures to characterize the unfolding of the native state. The first signature is based not on the geometrical model described below, but is calculated directly from the crystallographic data. We denote the lateral distance between terminal alpha carbons in a n=3 triplet centered on the "midpoint" residue i as Tn(i). The lateral distance between terminal residues in two (or more) adjacent triplets, connected in a linear array, can also be calculated directly from the data. We denote the lateral distance between terminal alpha carbons in a modular unit of two, three, .n triplets, each centered on the "midpoint" residue i, by T(i). The ratio T(i)/Tn(i) then gives a measure of the lateral, linear extension of the polypeptide chair in a given stage of unfolding relative to the native state.

For any given segment of the polypeptide chain, a sum of the T(i)/Tn(i) for all members can be calculated, and an overall average constructed. For example, for the 9 residues in the helical segment closest to the N-terminal end of intelectin-1, SER 40 to CYS 48, we calculate

$$< L > = \quad [\ T(40)/Tn(40)\ + \ldots\ + T(48)/Tn(48)\ ]/9 = 1.98.$$

If there were no steric interference among the side chains of these 9 residues, the value of $< L >$ would be exactly 2. As is seen in Table 1, the values of $<L>$ for each stage of extension of this nine-residue segment are close to (but not exactly equal to) integers, as is the case for the other helical regions in intelectin-1.

The two remaining signatures of our geometrical approach can be illustrated using Fig.1. There, the nine residue, helical segment, PDB 40 to 48, is displayed in a fully-extended, linear array of four triplets. We are interested in calculating the change in position of the center-most residue in this segment, ILE 44, as the native state unfolds to this extended state. A calcium ion (PDB 401) is assigned as the origin of a coordinate system. Choice of either of the other Ca ions as origin has a minimal effect on the quantitative results, and no effect on the qualitative conclusions drawn from the data.

To determine the positional change of ILE 44, we proceed as follows. First, the locations of the alpha-carbons of the residues SER 40 and CYS 48 are determined from the crystallographic data. The linear distance D(R40-R48) between the two terminal alpha-carbons in the native state is calculated using the Theorem of Pythagoras. Inter-residue distances, determined from the crystallographic data, are used in conjunction with the Law of Cosines to determine the angles α, β and γ. Finally, the linear distance T(R40-R48) is used in conjunction with the Law of Sines, to determine the displacement f(44) of the central residue (ILE 44) from the Ca (PDB 401).

A similar procedure is followed for all nine residues in the helical segment (SER 40 to CYS 48), and an overall average calculated:

$$\langle f \rangle = [\, f(40) + \ldots \; f(48) \,]/\, 9 = 1.34.$$

Values of this signature for this helical segment, as well as for the other helical regions in intelectin-1, are given in Table 1 for six stages of unfolding of the polypeptide chain.

Finally, a signature for the "spread" in angle as one goes from the native state to the third

extended state in the above example (n=9, three adjacent triplets in a linear array) can be calculated as follows. First, for the native state we construct the sum

$$\Sigma\,(n=3) = [\,\beta(39 \text{ to } 41) + \ldots + \beta(47 \text{ to } 49)\,]/9,$$

Similarly, for the extended state, we determine

$$\Sigma(n=9) = [\,\beta(36 \text{ to } 44) + \ldots + \beta(44 \text{ to } 52)\,]/9.$$

We define the difference between these two measures,

$$< \beta^{o} > = \Sigma(n=9) - \Sigma(n=3)$$

and find that for the 9-residue helix (SER 40 to CYS 48), $< \beta^{o} > = 9.54^{o.}$. Values of the lateral extension < L >, the radial extension < f >, and the angular "spread" $< \beta^{o} >$ are calculated for each residue of the protein, and for each stage of unfolding.

Two technical features of our approach need mention. First, owing to the choice of the triplet as the modular unit in our calculations, calculations of $< f_n >$ and $< \beta_n{}^{o} >$ for n-residue segments in the near vicinity of the N-terminal and C-terminal end of the polypeptide chain are (somewhat) less reliable than for n-residue segments interior to the polypeptide chain. In fact, our analysis cannot even be applied to two residues, the N-terminal residue and the C-terminal residue.

Second, one of the diagnostics used to assess the crystallographic data describing a given protein is the α-carbon to α-carbon distance between adjacent residues which, theoretically, is 3.80 Å. This theoretical constraint is taken into account in reporting crystallographic data. In the perturbed states that result from application of our geometrical model, nearest-

neighbor α-carbon to α-carbon distances can vary slightly from the theoretical value, 3.80 Å. Here, as in Ref(4), rather than adjusting the individual residue-to-residue distances to be 3.80 Å, we impose a statistical constraint: the overall average α-carbon to α-carbon distance for all nearest-neighbor residues in the polypeptide chain at each stage is adjusted to be 3.80 Å . Differences between values of $< f >$ and $< \beta^o >$ calculated with or without this adjustment are significant only for the most extended configurations.

Apart from the above two caveats, no further approximations are introduced in calculating $< f_n >$ and $< \beta_n{}^o >$ for n-residue segments. In our deterministic model, the underlying theory is based on Euclidean geometry and the only mathematical functions needed are sines and cosines. On a standard workstation, the longest "compilation time" required to calculate the signatures reported here is the time it takes to enter the crystallographic data for a given protein.

In the following sections we present and discuss the results on chain A of intelection-1 generated using our approach, first globally for the protein as a whole, then locally for the 12 helices in chain A, and finally for residues in the vicinity of the binding site.

II. Global Results

We present in Fig. 2 the signatures, $< L >$, $< f >$ and $< \beta^o >$ for all residues of intelectin-1 (chain A) in the first stage of extension, that is, modular units of n =5 residues, a linear array of two connected triplets. Using the regional specifications reported in this reference,[5] the figures are color coded as follows. Alpha helices are in blue, the sheets in red, SS bonds in yellow, and the remaining residues in green. Similar plots have been constructed for

all six stages of unfolding, but already the main features of destabilization of the native structure are displayed in these figures.

As illustrated in Fig. 2a, values of the lateral extension signature Tn for helical residues are systematically smaller than for non-helical regions or for the sheets. This behavior is also evident in values of the radial signature f displayed in Fig. 2b. The profile of angular "spread" versus residue, Fig. 2c, reveals that changes in helical regions are less pronounced than in non-helical ones.

Corroborative evidence presented in Table 2 reveals that values of < L > for the protein as a whole are (or nearly) integers. For example, the overall value of <Tn> for the native state is 6.159, and the value calculated for the first extended state is 12.312, hence the value 2.00 for the ratio given in the table.

Values of < f> are essentially constant in two regimes. In the first three stages of destabilization, < f > is 1.37, whereas in the latter three stages the value is 1.41 (+/- 0.01). The "break point" follows the extension n=9, which corresponds to the disruption of hydrogen bonds in two convolutions of an α-helix. The simplest interpretation of these data is that the maximum length of helices for chain A of intelectin-1 is nine residues. Hence, once the modular unit in our model exceeds nine residues, all segments "behave like" non-helical regions.

With increasing destabilization, values of the angular "spread" show a gradual decrease in angular difference between the extended state and the native state. This behavior can be traced back to the more significant changes in the non-helical regions (versus the helical regions) in the early stages of unfolding. These changes become less significant as longer segments

of the polypeptide chain unfold.

A visualization of the global behavior of chain A of intelectin-1 is displayed in Fig. 3. Changes in the signatures <L>, <f> and <β^o> are not independent, but occur synchronously. Fig. 3 gives a sense of the overall change in structure of the native state as it evolves to the first unfolded state.

II. Helical Regions

We return to the data given in Table 1 for the 12 helical regions in chain A of intelectin-1. To interpret these regional data, we use as a guide the global results on $< f >$ and $< \beta^o >$ reported in the previous section. There, we reported that in the first three stages of destabilization, values of $< f >$ for the protein as a whole were constant and centered on the value 1.37, whereas for the latter three stages, values centered on 1.41.

Using this metric, the 12 helices can be divided into two groups. In the first group, values of $< f >$ for all extensions are systematically less than 1.37. In this group are helices AA2 (5 residues), AA3 (5 residues), AA10 (6 residues). In the second group are helices that behave, more or less, "globally," i.e. values less than or equal to 1.37 for the first three stages and greater than 1.37 for the latter stages. In this group are helices AA1 (9 residues), AA4 (6 residues), AA6 (6 residues), AA7(5 residues), AA8 (9 residues), AA9(8 residues), AA11(5 residues) and AA12 (6 residues). As anticipated from our discussion of the global behavior, once the H-bonds in the longer helices (8 or 9 residues) of intelectin-1 are disrupted, the unfolding behavior should be similar to non-helical regions. The behavior of helices having 5 or 6 residues, i.e, those having a single convolution of an α-helix, is not uniform, and may

reflect environmental influences, e.g. the presence of nearby sheets or SS bonds.

Finally, there is a systematic increase in values of the angular signature $< \beta^o >$ for all helices as a given segment becomes destabilized. Here, we ask whether the angular "spread" for a given helix is about equal to the maximum value of $< \beta^o >$ calculated for the protein as a whole, viz., $< \beta^o > = 38^o$, or greater. The helices fall into two groups.

In the first group, where $< \beta^o >_{max}$ is less than 38^o, are helices AA1(9 residues), AA2(5 residues), AA8(9 residues), AA9(8 residues), AA10 (6 residues). Helices for which the angular "spread" is greater than $< \beta^o >_{max}$, are AA3(5 residues), AA4(6 residues), AA6(6 residues), AA7(5 residues), AA1(5 residues) and AA12(6 residues). We conclude that longer helices (with two convolutions) tend to be relatively more resilient to disruption of the angular configuration defining the native state.

Distinctions between breaking one, two, or more hydrogen bonds in α-helices has been the subject of an extensive series of theoretical and computational studies by Buehler and his colleagues. [24] In their generalization of the Bell model, and complementary molecular dynamics simulations, they study the breaking of hydrogen bonds between one, two, … convolutions of the alpha-helix. They identify a first transition state, after which further deformations lead to a second potential barrier (and transition state) at which further bond breakage occurs, and so on. Their generalization of the Bell model is in the same spirit as our use of n-residue, linear extensions of the polypeptide backbone to track the unfolding of the

protein. The two geometrical approaches give results that are in qualitative agreement.

IV. The Binding Site

Kiessling et al. highlight[5] the importance of the segment, residues 243 to 297, in the vicinity of the Ca ions and the binding site; see their Fig. 3. In Table 3a and in Fig. 4a and Fig. 4b we present our characterization of this region in terms of the three signatures, L, f and β^o. Comparison with the results presented for intelectin-1 as a whole, and the results for the helices in the previous section, show that the residues in this region are amazingly resilient to radial displacement and angular "spreading" through the six stages of destabilization considered in our study. This resiliency is also reflected in the graphical display of the native state and the first extended state, Fig. 5. The latter figure shows clearly the position of the three Ca ions identified in the structure, PDB 401 (in red), PDB 401 (in green) and PDB 403 (in blue). The (relatively) small changes in the unfolding signatures, also apparent from Fig. 5, show that the geometry of the protein in the vicinity of the binding site, when subject to geometric perturbations, is sensibly "conserved."

Kiessling et al. identify eight residues that bind to the three Ca ions.[5] Two are uncharged residues, ASN 243 and ASN 260; two are hydrophobic residues, TRP 288 and TYR 297; and four are charged residues, GLU 244, GLU 262, HIS 263, and GLU 274. Our calculations show that these residues contribute significantly to the resiliency of intelectin-1 to structural perturbations in the vicinity of the binding site. Radial displacement and angular "spread" from the native state are dramatically suppressed in the vicinity of these residues. Particularly striking is the near "freezing" of angular change in the vicinity of these residues as the protein unfolds from the native state.

The above assignments are supported by results reported in a study of the relative calcium binding strengths of amino acids.[25] The following ordering was found for 18 amino acids (data for ASP and GLU were not reported):

CYS < SER < THR < ILE < LEU < VAL < GLY < ALA < PRO <

PHE < MET < TYR < ASN < HIS < GLN < TRP < LYS < ARG

Four of the seven residues that contribute most to Ca binding are among the residues identified in this reference.[5]

To set the above results and conclusions in a broader context, we explore whether metal ions in other proteins we have studied also conferred structural stability on segments of the polypeptide chain. We report here on the three cytochromes, cytochrome c, cytochrome b_{562} and cytochrome c′ studied in this reference.[4]

To carry out our analysis, we scan the PDB data set for each cytochrome and select segments (of five or more residues) that are within 10 Angstroms of the iron atom. We then calculate values of $\langle f \rangle$ and $\langle \beta^o \rangle$ for each of the six stages of destabilization. As in Table 3a, we include for each entry the "all residue" average at each state; this average includes the segments being considered. Results for the three cytochromes are presented in Tables 3b, 3c and 3d.

Interesting comparisons can be drawn between and among intelectin and the cytochromes, but the overarching conclusion is that one metal ion (iron) in the cytochromes and,

for sure, the three calcium ions in intelectin, play a fundamental role in stabilizing the polypeptide chain in the near vicinity of the ion(s), at least in the early stages of unfolding. Importantly, these data and the results presented earlier in this section show that, owing to the presence of three calcium ions, the long chain surrounding the binding site in intelectin is amazingly resilient to structural perturbations.

V. Discussion

Recent studies suggest that intelectin-1 is an antimicrobial protein.[5] Earlier work had drawn attention to its role in mesothelioma[26] and gastric cancer.[27] The present study has been carried out to explore systematically the resiliency of intelectin-1 (chain A) to structural perturbations, particularly as they influence the stability of helices and residues in the vicinity of the binding site. We have quantified the destabilization through six stages of unfolding of the polypeptide chain.

The global behavior of the protein is described in Section II, the behavior of the several helices in Section III and behavior in the vicinity of the binding site in the previous section. The "personality" of the protein in these regions is distinguished by three signatures determined from the crystallographic data presented in this reference,[5] and elaborated using our geometrical model. We stress that once the model is defined, the consequences are developed without any further approximations.

Quantitative differences are established between the global behavior of the lateral, radial and angular signatures, and the (typically smaller) values calculated for the helices and

residues in the vicinity of the binding site. If these local regions tend to be more resilient to destabilization, there must be other regions of intelectin-1, where more dramatic changes occur, thus resulting in the overall averages reported in Table 2. The visual evidence presented in Fig. 2 shows clearly that changes in the signatures <f> and <β^o> appear to be more pronounced for residues < 100. In this region are located three sheets. As quantified in Table 4, values of the radial signature <f> for the sheets defined by residues GLY 54 to ARG 59, ILE 65 to ASP 71, and TRP 79 to VAL 85, are significantly larger than the values calculated for the α-helices, or for residues in the vicinity of the binding site.

Note that for all three sheets, the value for the first two stages of extension is <f> = 1.5 but then jumps to <f> =1.6 once the third extension is considered. Recalling the discussion of alpha-helices in Section III, we associate this "breakpoint" in quantitative behavior with the disruption of two H-bonds.

In closing, we have focused on chain A of intelectin-1 [4WMQ]. However, Kiessling et al. have since reported crystal structures for the associated proteins: 4WMQ (chain B), 4WMY (chain A and B), 4WN0 (one chain), and 4WMO (chains A through F).[5,28] We have performed calculations complementary to those reported in this study for each of these (nine) proteins. A comprehensive analysis of factors influencing the stability of the native state of these lectin proteins will be presented subsequently.

Acknowledgements

We are delighted that our work is included in the special issue of Journal of Physical Chemistry in honor of Professor Mark Gordon, an all-around good guy who has made groundbreaking contributions to the quantum chemistry of atoms and molecules. His development of (the computational program) GAMESS has had a big impact on the field. One of the authors (JJK) is indebted to Dr. K. Wangkanont for clarifying details of the crystal structure of the lectin proteins cited in the Discussion. We thank the National Institutes of Health (R01 DK 019038 to HBG) for support of work performed at the California Institute of Technology. Financial support for R.A.G.-L. was provided by the Howard Hughes Medical Institute Research Program from Pomona College. The molecular graphics images were produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41 RR-01081).

References

(1) Warren, J. J. ; Gray, H. B.; Winkler, J. R.; Kozak, J. J. A Euclidean Perspective on the Unfolding of Azurin: Spatial Correlations. *Mol. Phys.* **2013,** *111*, 922-929.

(2) Warren, J. J.; Gray, H. B.; Winkler, J. R.; Kozak, J. J. Euclidean Perspective on the Unfolding of Azurin: Angular Correlations. *Mol. Phys.* **2013,** *111*, 3762-3769.

(3) Gray, H. B.; Warren, J. J.; Winkler, J. R.; Kozak, J. J. A Euclidean Perspective on the Unfolding of Azurin: Chain Motion. *J. Biol. Inorg. Chem.* **2014,** *19*, 555-563.

(4) Kozak, J. J.; Gray, H. B.; Garza-López, R. A. Cytochrome Unfolding Pathways from Computational Analysis of Crystal Structures. *J. Inorg. Biochem.* **2016,** *155*, 44-55.

(5) Wesener, D. A.; Wangkanont, K.; McBride, R.; Song, X. Z.; Kraft, M. B.; Hodges, H. L.; Zarling, L. C.; Splain, R. A.; Smith, D. F.; Cummings, R. D.; et.al. Recognition of Microbial Glycans by Human Intelectin-1. *Nat. Struct. Mol. Biol.* **2015**, *22,* 603-610.

(6) Wanganot, K.; Kiessling ,L.L.; Forest, K.T. Full wwPDB X-ray Structure Validation Report, Structure of Human Intelectin-1 *PDB,***2015**, 1-12.

(7) Winkler, J. R.; Wittung-Stafshede, P.; Leckner, J.; Malmstrom, B. G.; Gray, H. B. Effects of Folding on Metalloprotein Active Sites. *P. Natl. Acad. Sci. USA* **1997,** *94*, 4246-4249.

(8). Liu, H.; Wang, C. P.; Chen, F.; Shen, S. H. Proteomic Analysis of Oil Bodies in Mature Jatropha Curcas Seeds with Different Lipid Content. *J. Proteomics* **2015,** *113*, 403-414.

(9) Scaramello, C. B. V.; Muzi, H.; Zapata-Sudo, G.; Sudo, R. T.; Cunha, V. D. N. FKBP12 Depletion Leads to Loss of Sarcoplasmic Reticulum Ca^{2+} Stores in Rat Vas Deferens. *J Pharmacol. Sci.* **2009,** *109*, 185-192.

(10) Zimmer, D. B.; Chaplin, J.; Baldwin, A.; Rast, M. S100-Mediated Signal Transduction in the Nervous System and Neurological Diseases. *Cell. Mol. Biol.* **2005,** *51*, 201-214.

(11) Biekofsky, R. R.; Turjanski, A. G.; Estrin, D. A.; Feeney, J.; Pastore, A. Ab Initio Study of NMR N-15 Chemical Shift Differences Induced by Ca^{2+} Binding to EF-Hand Proteins. *Biochem.* **2004,** *43*, 6554-6564.

(12) Kataeva, I. A.; Uversky, V. N.; Ljungdahl, L. G. Calcium and Domain Interactions Contribute to the Thermostability of Domains of the Multimodular Cellobiohydrolase, CbhA, a Subunit of the Clostridium Thermocellum Cellulosome. *Biochem. J.* **2003,** *372*, 151-161.

(13) Deloulme, J. C.; Gentil, E. J.; Baudier, J. Monitoring of S100 Homodimerization and Heterodimeric Interactions by the Yeast Two-Hybrid System. *Microsc. Res. Techniq.* **2003,** *60*, 560-568.

(14) Ahvazi, B.; Boeshans, K. M.; Idler, W.; Baxa, U.; Steinert, P. M. Roles of Calcium Ions in the Activation and Activity of the Transglutaminase 3 Enzyme. *J. Biol. Chem.* **2003**, *278*, 23834-23841.

(15) Jaren, O. R.; Kranz, J. K.; Sorensen, B. R.; Wand, A. J.; Shea, M. A. Calcium-Induced Conformational Switching of Paramecium Calmodulin Provides Evidence for Domain Coupling. *Biochem.* **2002,** *41*, 14158-14166.

(16) Koshiba, T.; Kobashigawa, Y.; Demura, M.; Nitta, K. Energetics of Three-State Unfolding of a Protein: Canine Milk Lysozyme. *Protein Eng.* **2001,** *14*, 967-974.

(17) Heath, D. J.; Downes, S.; Verderio, E.; Griffin, M. Characterization of Tissue Transglutaminase in Human Osteoblast-Like Cells. *J. Bone Miner. Res.* **2001,** *16*, 1477-1485.

(18) Prodhom, B.; Karplus, M. The Nature of the Ion-Binding Interactions in Ef-Hand Peptide Analogs- Free-Energy Simulation of Asp to Asn Mutations. *Protein Eng.* **1993**, *6*, 585-592.

(19) Perronebizzozero, N. I.; Cansino, V. V.; Kohn, D. T. Posttranscriptional Regulation of Gap-43 Gene-Expression in Pc12 Cells through Protein-Kinase-C Dependent Stabilization of the Messenger-Rna. *J. Cell Biol.* **1993,** *120*, 1263-1270.

(20) Kalabokis, V. N.; Santoro, M. M.; Hardwicke, P. M. D. Effect of Na^{+} and Nucleotide on the Stability of Solubilized Ca^{2+}-Free Ca-Atpase from Scallop Sarcoplasmic-Reticulum. *Biochem.* **1993,** *32*, 4389-4396.

(21) Bryan, P.; Alexander, P.; Strausberg, S.; Schwarz, F.; Lan, W.; Gilliland, G.; Gallagher, D. T. Energetics of Folding Subtilisin Bpn'. *Biochem.* **1992,** *31*, 4937-4945.

(22) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of Polypeptide Chain Configurations. *J. Mol. Biol.* **1963,** *7*, 95-99.

(23) Kozak, J. J.; Piasecki, J.; Szymczak, P. Distribution Function Approach to the Stability of Fluids, *Adv. Chem. Phys.* **2016** , *162*, 359-394.

(24) Bertaud, J.; Hester, J.; Jimenez, D. D.; Buehler, M. J. Energy Landscape, Structure and Rate Effects on Strength Properties of Alpha-Helical Proteins. *J. Phys-Condens. Mat.* **2010**, *22*, 1-14.

(25) Ho, Y. P.; Yang, M. W.; Chen, L. T.; Yang, Y. C. Relative Calcium-Binding Strengths of Amino Acids Determined Using the Kinetic Method. *Rapid Commun. Mass. Spectrom.* **2007,** *21*, 1083-1089.

(26) Tsuji, S.; Tsuura, Y.; Morohoshi, T.; Shinohara, T.; Oshita, F.; Yamada, K.; Kameda, Y.; Ohtsu, T.; Nakamura, Y.; Miyagi, Y. Secretion of Intelectin-1 from Malignant Pleural Mesothelioma into Pleural Effusion. *Brit. J. Cancer* **2010,** *103*, 517-523.

(27) Li, D.; Zhao, X.; Xiao, Y.; Mei, H.; Pu, J. R.; Xiang, X.; Jiao, W. J.; Song, H. J.; Qu, H. X.; Huang, K.; et.al. Intelectin 1 Suppresses Tumor Progression and is Associated With Improved Survival in Gastric Cancer. *Oncotarget* **2015,** *6*, 16168-16182.

(28) Wangkanont, K.; Wesener, D. A.; Vidani, J. A.; Kiessling, L. L.; Forest, K. T. Structures of Xenopus Embryonic Epidermal Lectin Reveal a Conserved Mechanism of Microbial Glycan Recognition. *J. Biol. Chem.* **2016,** *291*, 5596-5610.

Table 1. Signatures <L>, <β^o> and <f> intelectin-1 (chain A) helical regions.

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA1 (9)	40-48	Native	3	1.0	1.0	0.0
		1	5	1.98	1.20	- 0.10
		2	7	3.00	1.24	4.25
		3	9	4.03	1.34	9.54
		4	11	5.09	1.42	15.32
		5	13	6.15	1.46	19.99
		6	15	7.28	1.49	24.17

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA2 (5)	73-77	Native	3	1.0	1.0	0.0
		1	5	1.96	1.19	2.81
		2	7	2.98	1.03	4.21
		3	9	4.01	0.97	5.05
		4	11	5.05	1.00	8.63
		5	13	6.07	1.02	12.80
		6	15	7.19	1.15	13.63

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA3 (5)	125-129	Native	3	1.0	1.0	0.0
		1	5	2.01	1.32	13.74
		2	7	3.08	1.32	30.44
		3	9	4.21	1.23	36.71
		4	11	5.35	1.11	45.74
		5	13	6.50	1.12	44.16
		6	15	7.68	1.09	35.01

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA4 (6)	137-142	Native	3	1.0	1.0	0.0
		1	5	2.03	1.24	2.43
		2	7	3.11	1.38	10.83
		3	9	4.19	1.49	20.91
		4	11	5.33	1.56	36.44
		5	13	6.49	1.56	49.86
		6	15	7.64	1.54	56.35

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA5 (5)	157-165	Native	3	1.0	1.0	0.0
AA6 (6)		1	5	1.98	1.29	6 .88
		2	7	3.00	1.32	14.44
		3	9	4.02	1.32	20.80
		4	11	5.08	1.39	28.38
		5	13	6.16	1.30	26.01
		6	15	7.25	1.27	30.29

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA6 (6)	160-165	Native	3	1.0	1.0	0.0
		1	5	1.96	1.30	8.04
		2	7	3.00	1.32	15.80
		3	9	4.02	1.34	23.94
		4	11	5.12	1.52	41.82
		5	13	6.19	1.50	38.72
		6	15	7.31	1.53	43.84

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA7(5)	174-178	Native	3	1.0	1.0	0.0
		1	5	2.01	1.33	7.09
		2	7	3.05	1.40	13.05
		3	9	4.10	1.44	19.56
		4	11	5.25	1.50	26.17
		5	13	6.37	1.52	35.26
		6	15	7.52	1.55	43.40

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA8 (9)	182-190	Native	3	1.0	1.0	0.0
		1	5	1.99	1.18	4.81
		2	7	3.01	1.23	12.29
		3	9	4.04	1.32	19.16
		4	11	5.07	1.43	27.15
		5	13	6.12	1.50	31.92
		6	15	7.16	1.57	35.92

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA9 (8)	215-222	Native	3	1.0	1.0	0.0
		1	5	2.04	1.24	4.92
		2	7	3.09	1.37	14.01
		3	9	4.14	1.41	20.32
		4	11	5.21	1.50	27.30
		5	13	6.29	1.53	33.43
		6	15	7.39	1.58	38.38

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA10 (6)	224-229	Native	3	1.0	1.0	0.0
		1	5	2.04	1.28	7.51
		2	7	3.10	1.34	13.28
		3	9	4.17	1.25	16.18
		4	11	5.27	1.35	19.42
		5	13	6.37	1.26	19.31
		6	15	7.49	1.21	18.73

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA11 (5)	260-264	Native	3	1.0	1.0	0.0
		1	5	2.03	1.32	4.25
		2	7	3.14	1.38	7.01
		3	9	4.23	1.36	4.15
		4	11	5.44	1.40	9.17
		5	13	6.56	1.46	24.82
		6	15	7.79	1.51	41.67

HELIX (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA12 (6)	300-305	Native	3	1.0	1.0	0.0
		1	5	2.04	1.26	1.55
		2	7	3.12	1.37	17.94
		3	9	4.20	1.48	32.89
		4	11	5.32	1.57	44.84
		5	13	6.48	1.59	53.27
		6	15	7.61	1.64	62.73

Table 2. Global signatures <L>, <β^o> and <f> for intelectin-1, chain A.

Chain	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
A	all	Native	3	1.0	1.0	0.0
		1	5	2.00	1.37	10.08
		2	7	3.00	1.37	19.05
		3	9	3.99	1.37	25.44
		4	11	4.99	1.41	31.10
		5	13	5.99	1.40	35.02
		6	15	6.99	1.42	37.99

Table 3a. Signatures <L>, <f> and <β^o> for residues in vicinity of binding site of intelectin-1, chain A. All-residue averages in parentheses

Chain	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
A	243-297	Native	3	1.0	1.0	0.0
		1	5	2.02 (2.00)	1.34 (1.37)	8.71 (10.08)
		2	7	3.03 (3.00)	1.28 (1.37)	14.54 (19.05)
		3	9	4.06 (3.99)	1.25 (1.37)	16.88 (25.44)
		4	11	5.06 (4.99)	1.26 (1.41)	18.64 (31.10)
		5	13	6.09 (5.99)	1.25 (1.40)	20.78 (35.02)
		6	15	7.10 (6.99)	1.23 (1.42)	22.74 (37.99)

Table 3b. Signatures <f> and <β^o> for residues in near vicinity of iron ion in cytochrome c . All-residue averages in parentheses.

Extension	n-unit	Residues	<f>	<β^o>	Residues	<f>	<β^o>	Residues	<f>	<β^o>
Native	3	13-19	1.0	0.0	28-32	1.0	0.0	78-82	1.0	0.0
1	5		1.16	14.98		1.18	10.32		1.42	24.81
			(1.21)	(13.02)		(1.21)	(13.02)		(1.21)	(13.02)
2	7		1.17	27.69		1.05	29.14		1.31	41.04
			(1.17)	(26.53)		(1.17)	(26.53)		(1.17)	(26.53)
3	9		1.19	33.37		1.12	34.07		1.29	40.26
			(1.18)	(35.18)		(1.18)	(35.18)		(1.18)	(35.18)
4	11		1.23	36.00		1.15	29.29		1.22	32.70
			(1.19)	(41.90)		(1.19)	(41.90)		(1.19)	(41.90)
5	13		1.38	36.32		1.26	24.64		1.30	28.29
			(1.26)	(46.17)		(1.26)	(46.1)		(1.26)	(46.17)
6	15		1.25	37.80		1.20	27.08		1.17	26.60
			(1.18)	(51.26)		(1.18)	(51.26)		(1.18)	(51.26)

Table 3c. Signatures <f> and <β^o> for residues in near vicinity of iron ion in cytochrome b-562. All-residue averages in parentheses.

Extension	n-unit	Residues	<f>	<β^o>	Residues	<f>	<β^o>
Native	3	3-11	1.0	0.0	97-104	1.0	0.0
1	5	3-11	1.08 (1.15)	7.35 (4.69)	97-104	1.06 (1.15)	13.98 (4.69)
2	7	4-11	1.16 (1.17)	26.24 (16.35)	97-103	1.20 (1.17)	29.14 (41.56)
3	9	5-11	1.23 (1.23)	40.51 (24.61)	97-102	1.27 (1.23)	54.76 (24.61)
4	11	6-11	1.30 (1.23)	54.60 (33.15)	97-101	1.34 (1.23)	69.18 (33.15)
5	13	7-11	1.48 (1.39)	60.65 (40.92)	97-100	1.52 (1.39)	86.23 (40.92)
6	15	8-11	1.41 (1.25)	72.25 (48.30)	97-99	1.43 (1.25)	102.7 (48.30)

Table 3d. Signatures <f> and <β^o> for residues in near vicinity of iron ion in cytochrome c' . All-residue averages in parentheses.

Extension	n-unit	Residues	<f>	<β^o>	Residues	<f>	<β^o>
Native	3	8-16	1.0	0.0	112-121	1.0	0.0
1	5		1.11 (1.18)	8.30 (6.24)	112-121	1.11 (1.18)	13.08 (6.24)
2	7		1.16 (1.16)	28.05 (18.21)	112-120	1.14 (1.16)	35.27 (18.21)
3	9		1.14 (1.13)	41.09 (27.71)	112-119	1.11 (1.13)	51.37 (27.71)
4	11		1.30 (1.23)	53.72 (37.36)	112-118	1.27 (1.23)	67.73 (37.36)
5	13		1.36 (1.22)	66.08 (45.91)	112-117	1.31 (1.22)	74.75 (45.91)
6	15		1.38 (1.21)	76.67 (53.75)	112-116	1.36 (1.21)	86.43 (53.75)

Table 4. Signatures <L>, <β^o> and <f> intelectin-1 (chain A) sheets.

SHEET (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA1 (6)	54-59	Native	3	1.0	1.0	0.0
		1	5	1.98	1.50	12.81
		2	7	2.94	1.52	22.31
		3	9	3.88	1.56	27.67
		4	11	4.79	1.60	31.08
		5	13	5.69	1.60	29.99
		6	15	6.57	1.58	25.71

SHEET (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA2 (7)	65-71	Native	3	1.0	1.0	0.0
		1	5	1.99	1.51	14.03
		2	7	2.96	1.56	26.40
		3	9	3.93	1.60	34.94
		4	11	4.81	1.62	37.60
		5	13	5.77	1.61	37.52
		6	15	6.66	1.59	33.63

SHEET (#)	PDB Residues	Extension	n-unit	<L>	<f>	<β^o>
AA3 (7)	79-85	Native	3	1.0	1.0	0.0
		1	5	2.01	1.51	17.82
		2	7	3.01	1.57	32.59
		3	9	3.97	1.61	47.88
		4	11	4.97	1.66	59.39
		5	13	5.90	1.69	74.69
		6	15	6.86	1.71	91.41

Figure 1: Specification of the geometrical model for a nine residue segment in intelectin-1.

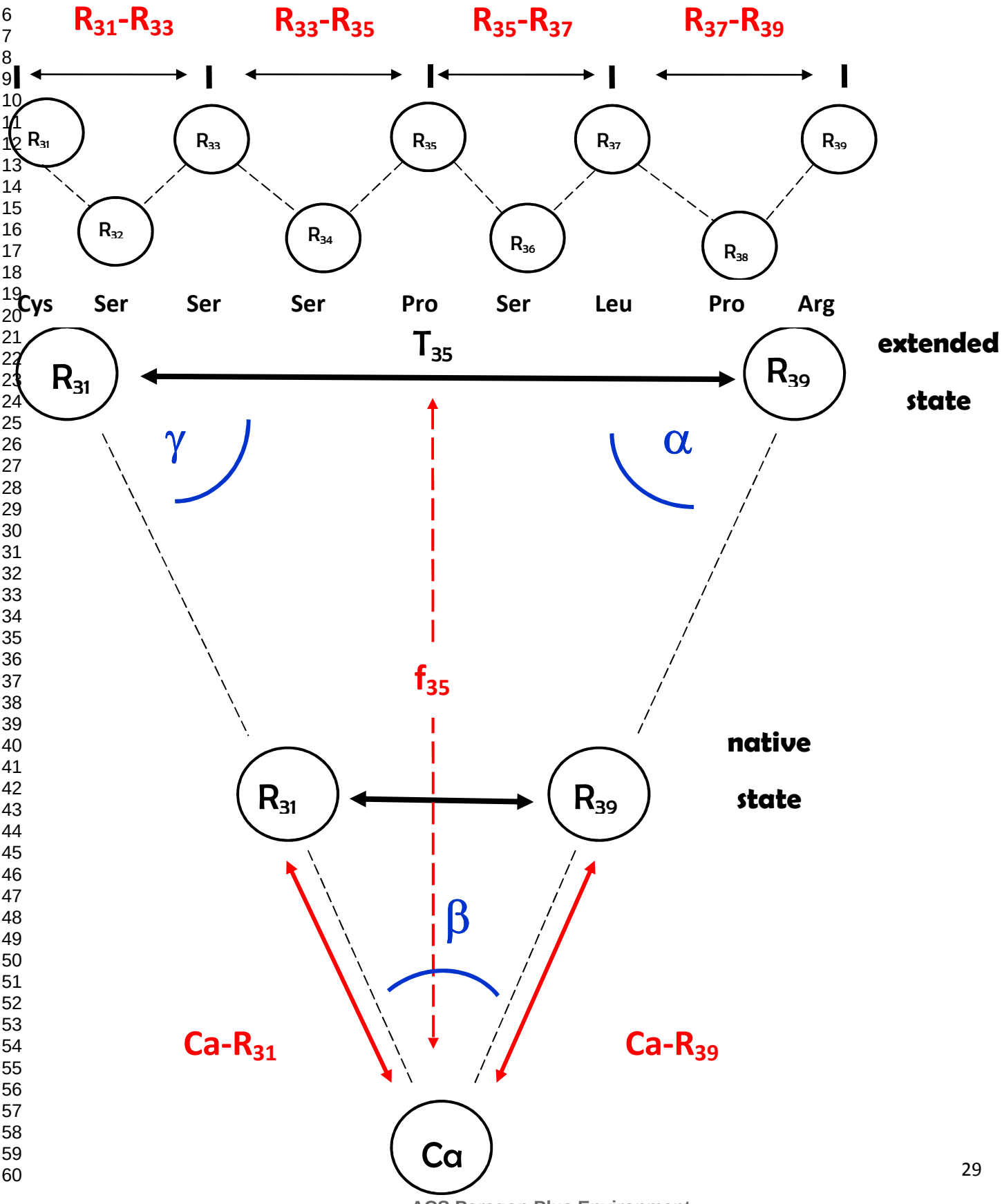


Figure 2a: Signature T vs PDB residue number for the first extended state for all residues. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

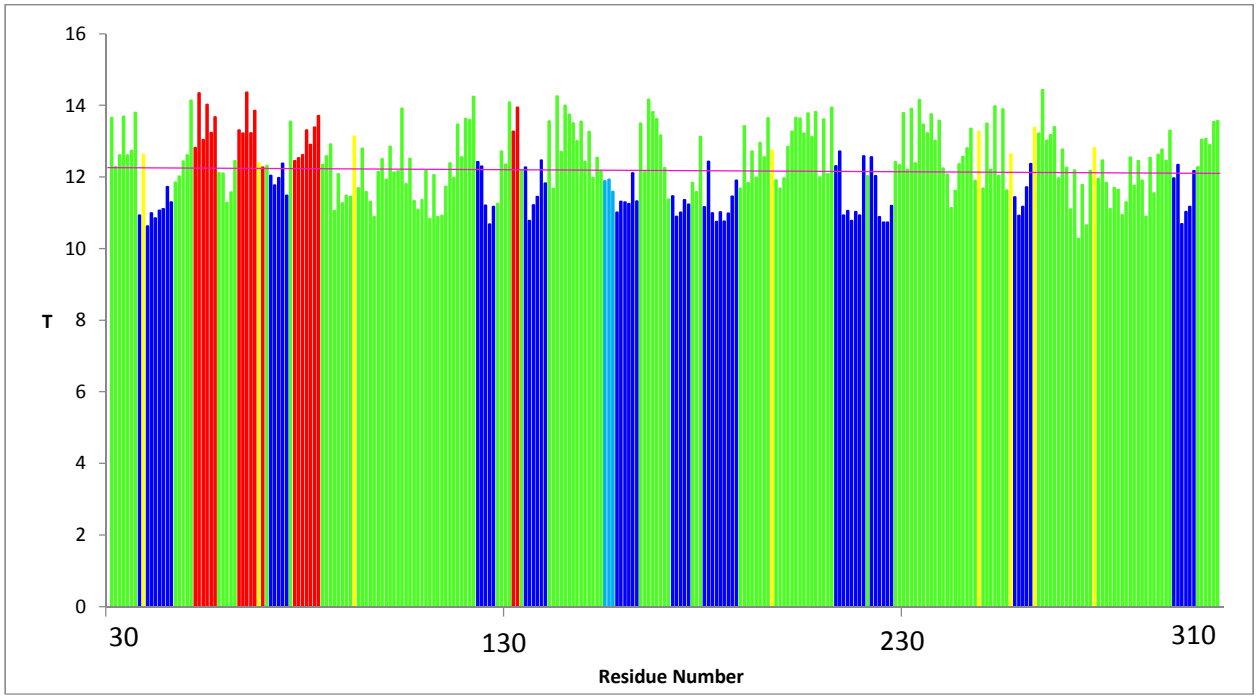


Figure 2b: Signature f vs PDB residue number for the first extended state for all residues. The orange line represents the native state, and the magenta line is the value of <f> for the region displayed. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

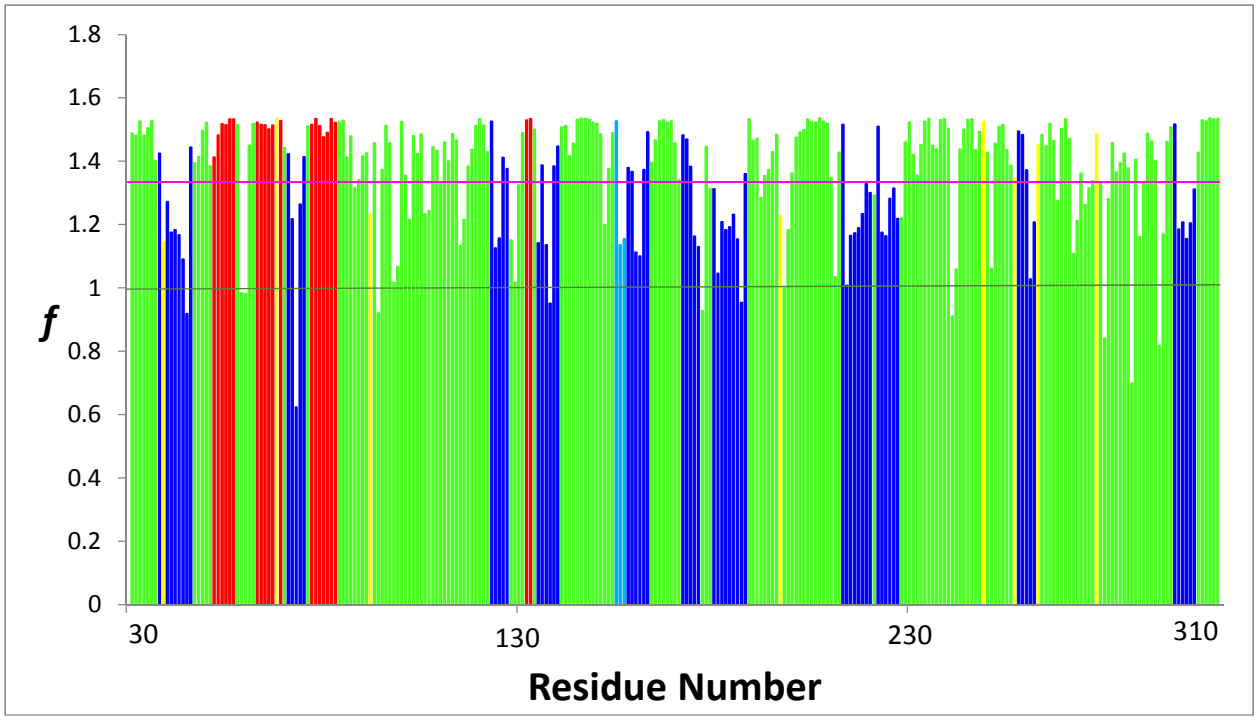


Figure 2c: Signature β vs PDB residue number for the first extended state for all residues. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

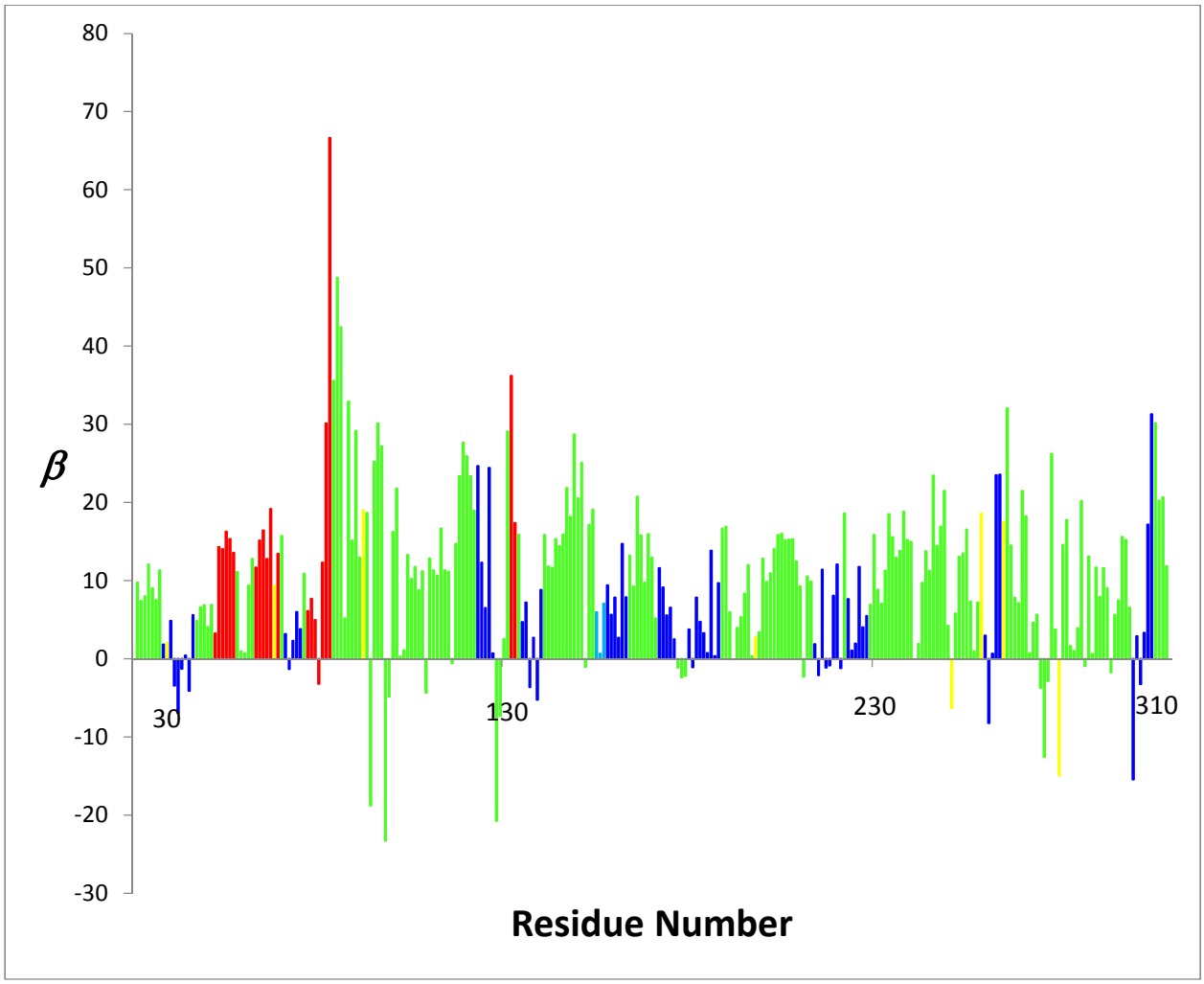


Figure 3: Chimera plot of the native (goldenrod) and first extended state (black) for the full protein. Calcium(II) ions in red, green and blue.

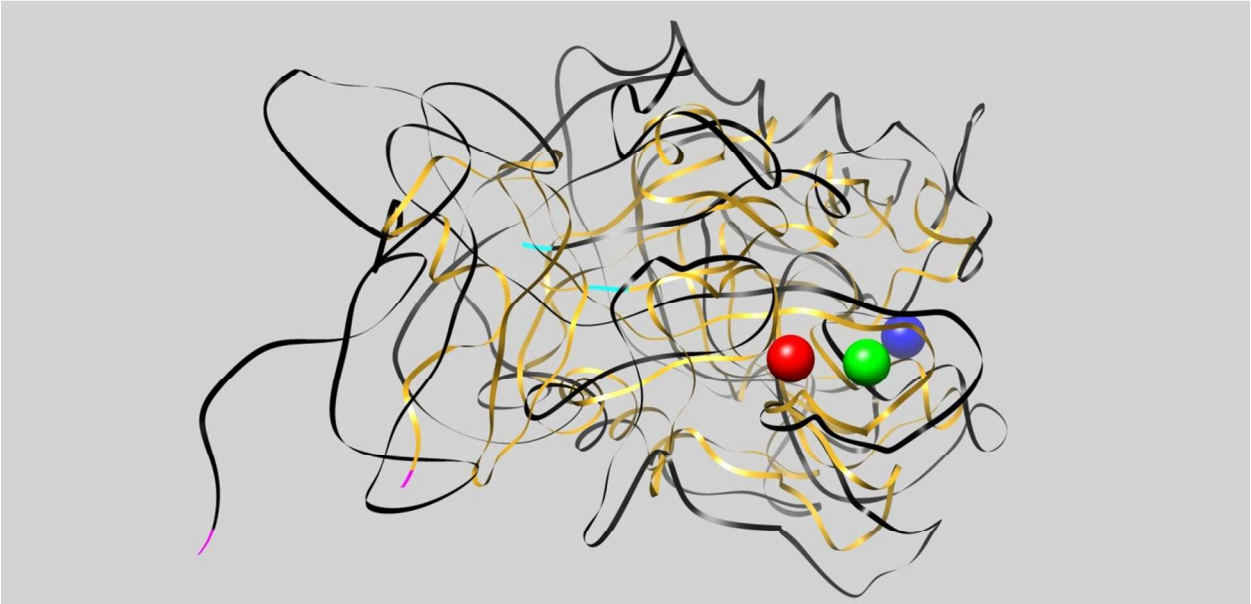


Figure 4a: Signature f vs PDB residue number for the first extended state for the (243-297) residues. The orange line represents the native state, and the magenta line is the value of <f> for the region displayed. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

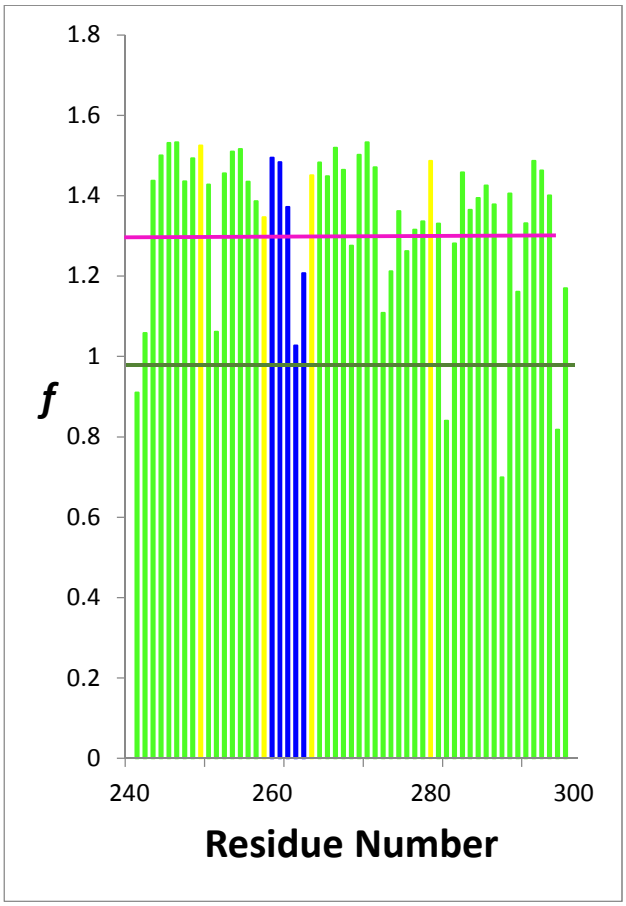


Figure 4b: Signature β vs PDB residue number for the first extended state for the (243-297) residues. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

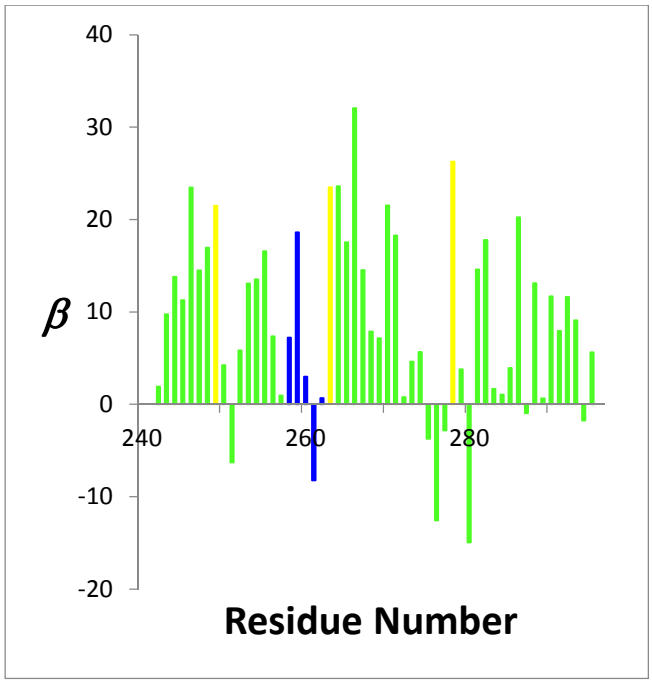
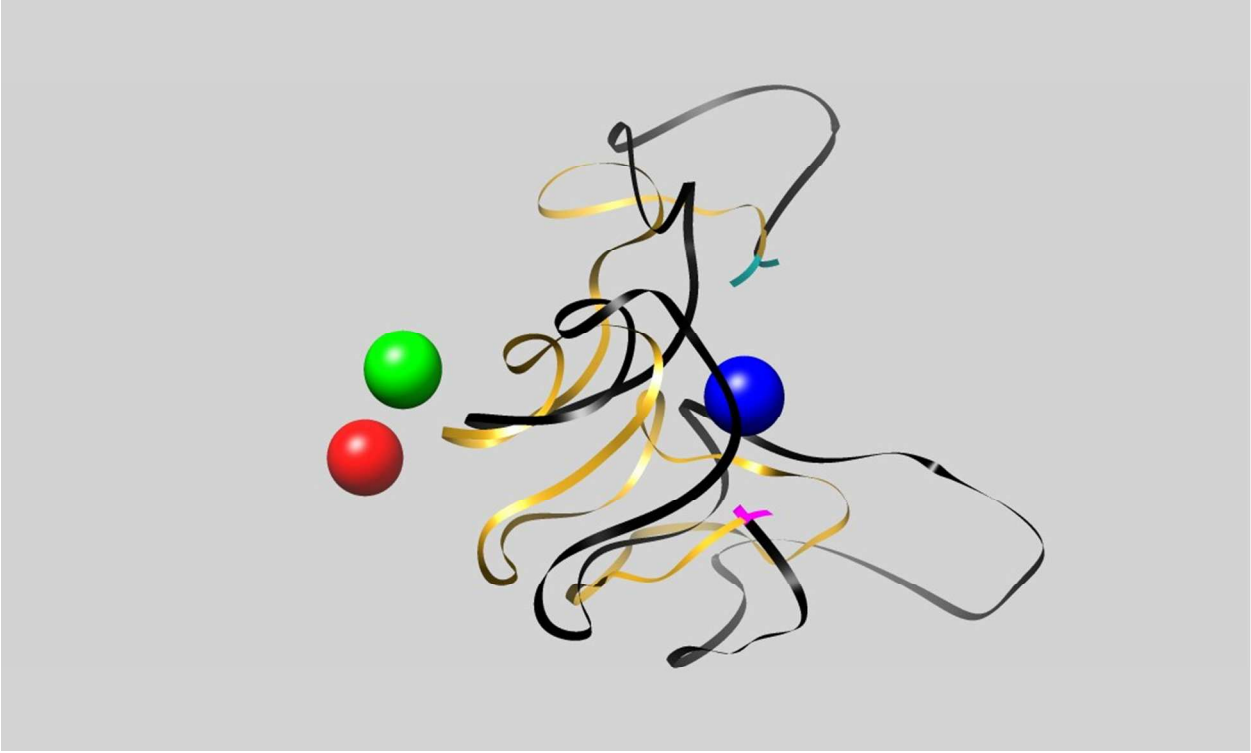


Figure 5: Chimera plot of the native (goldenrod) and first extended state (black) for the (243-297) residues. Calcium(II) ions in red, green and blue.



"Table of Contents Graphic"

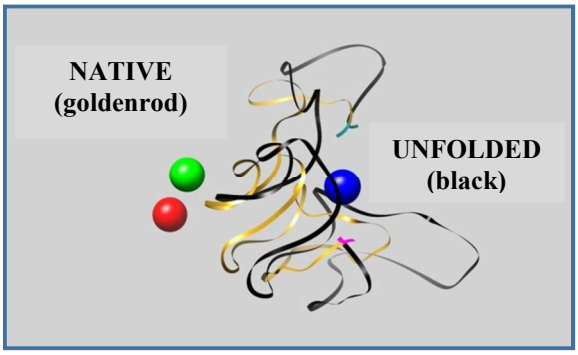


Figure 1: Specification of the geometrical model for a nine residue segment in intelectin-1.

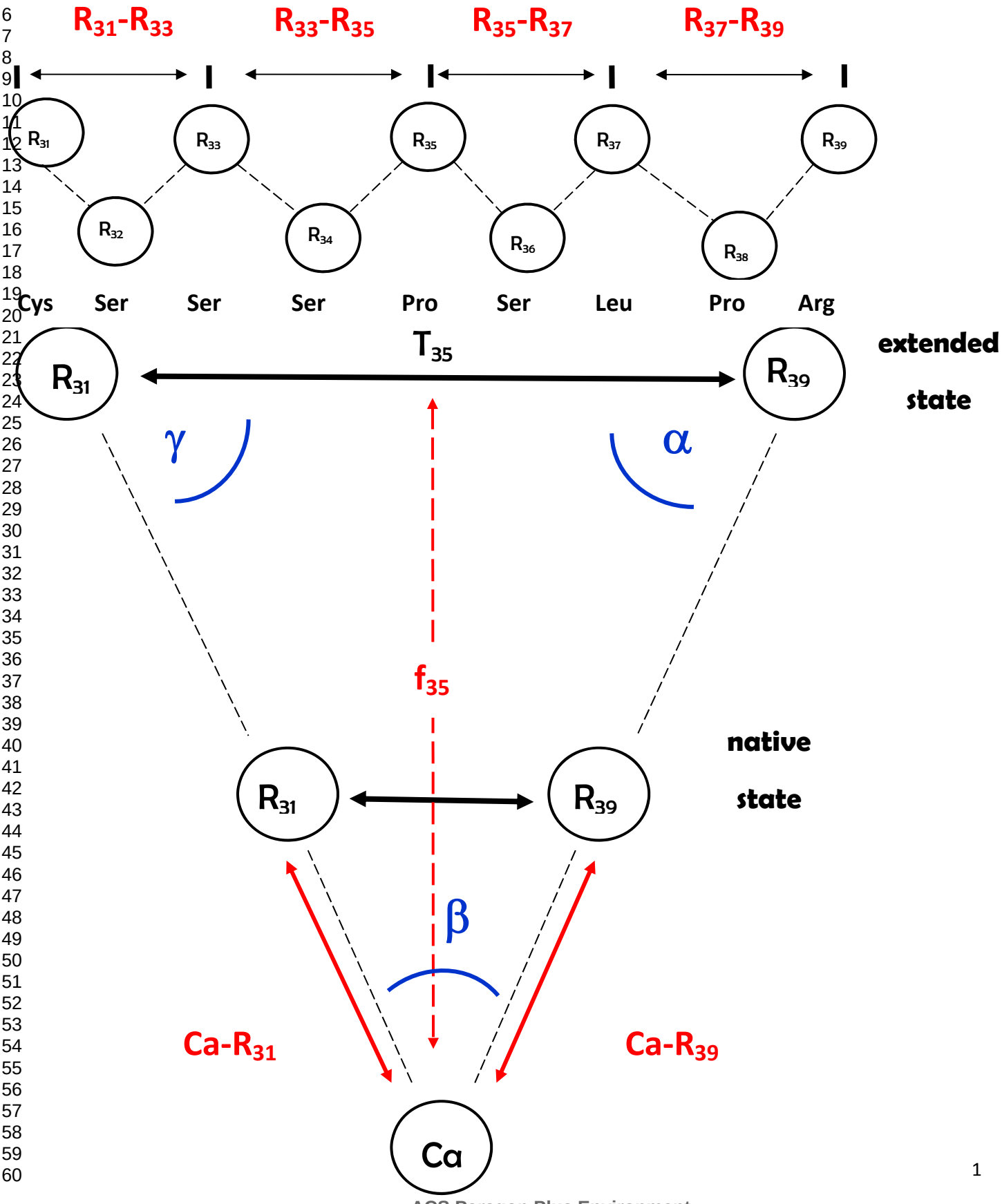


Figure 2a: Signature T vs PDB residue number for the first extended state for all residues. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

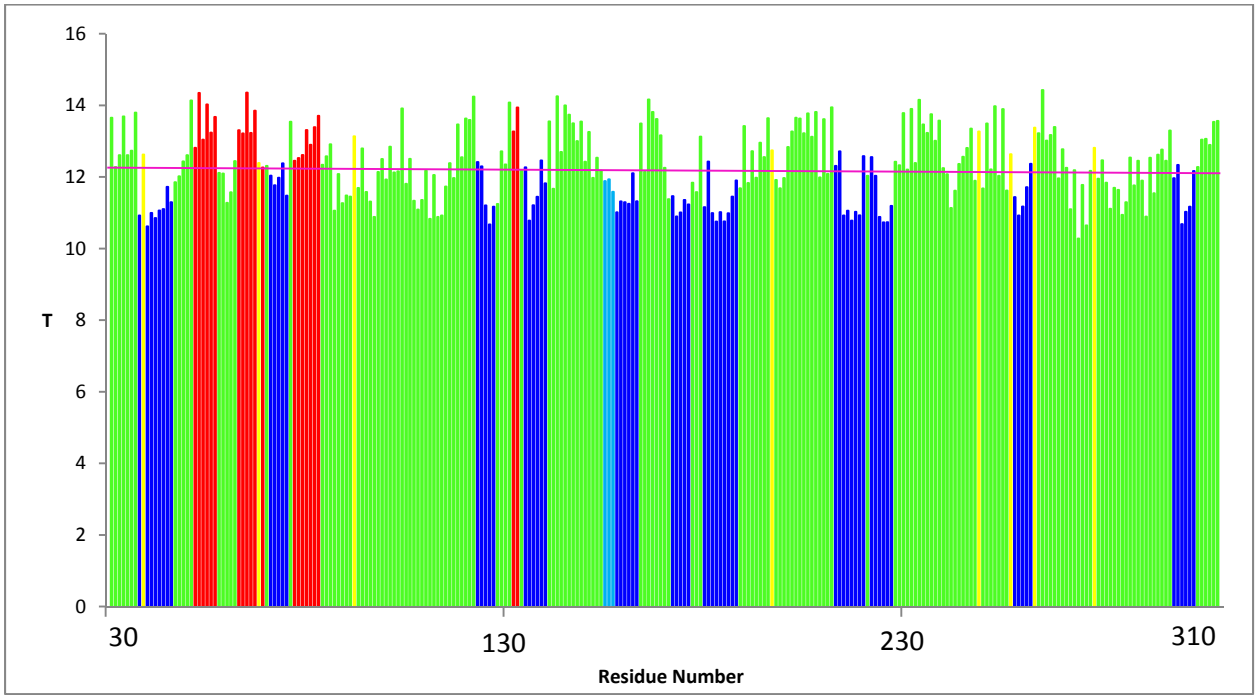


Figure 2b: Signature f vs PDB residue number for the first extended state for all residues. The orange line represents the native state, and the magenta line is the value of <f> for the region displayed. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

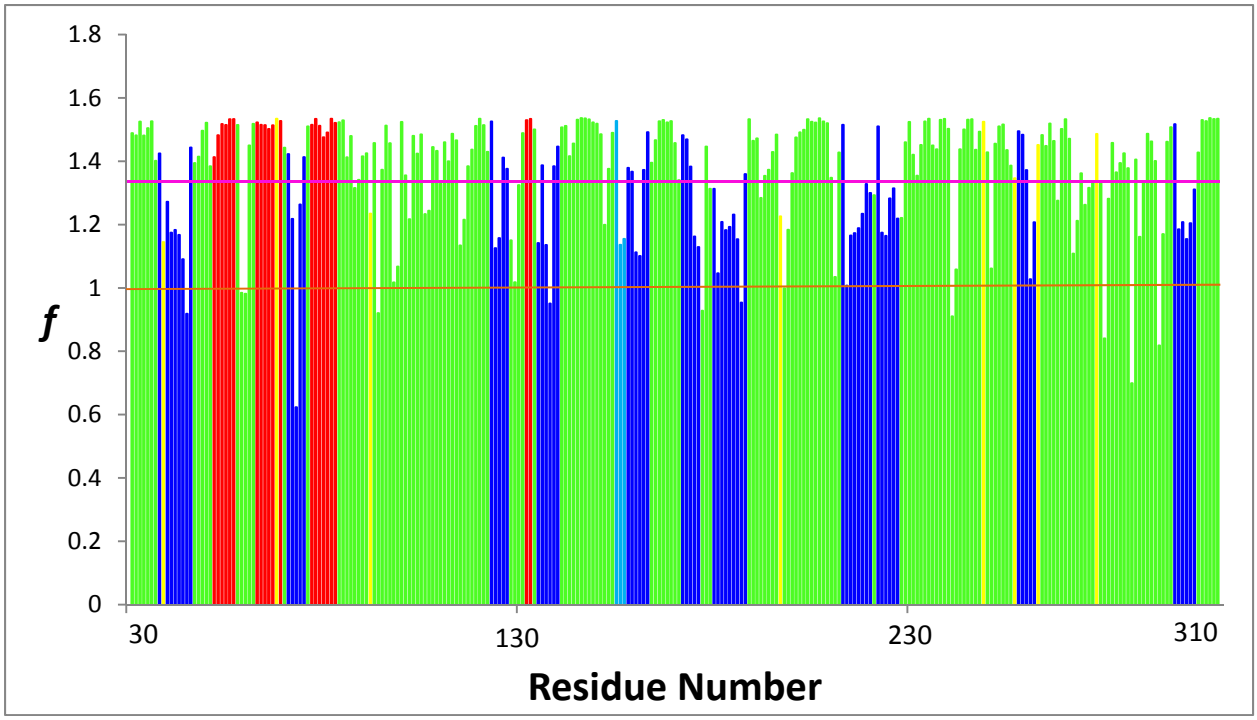


Figure 2c: Signature β vs PDB residue number for the first extended state for all residues. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

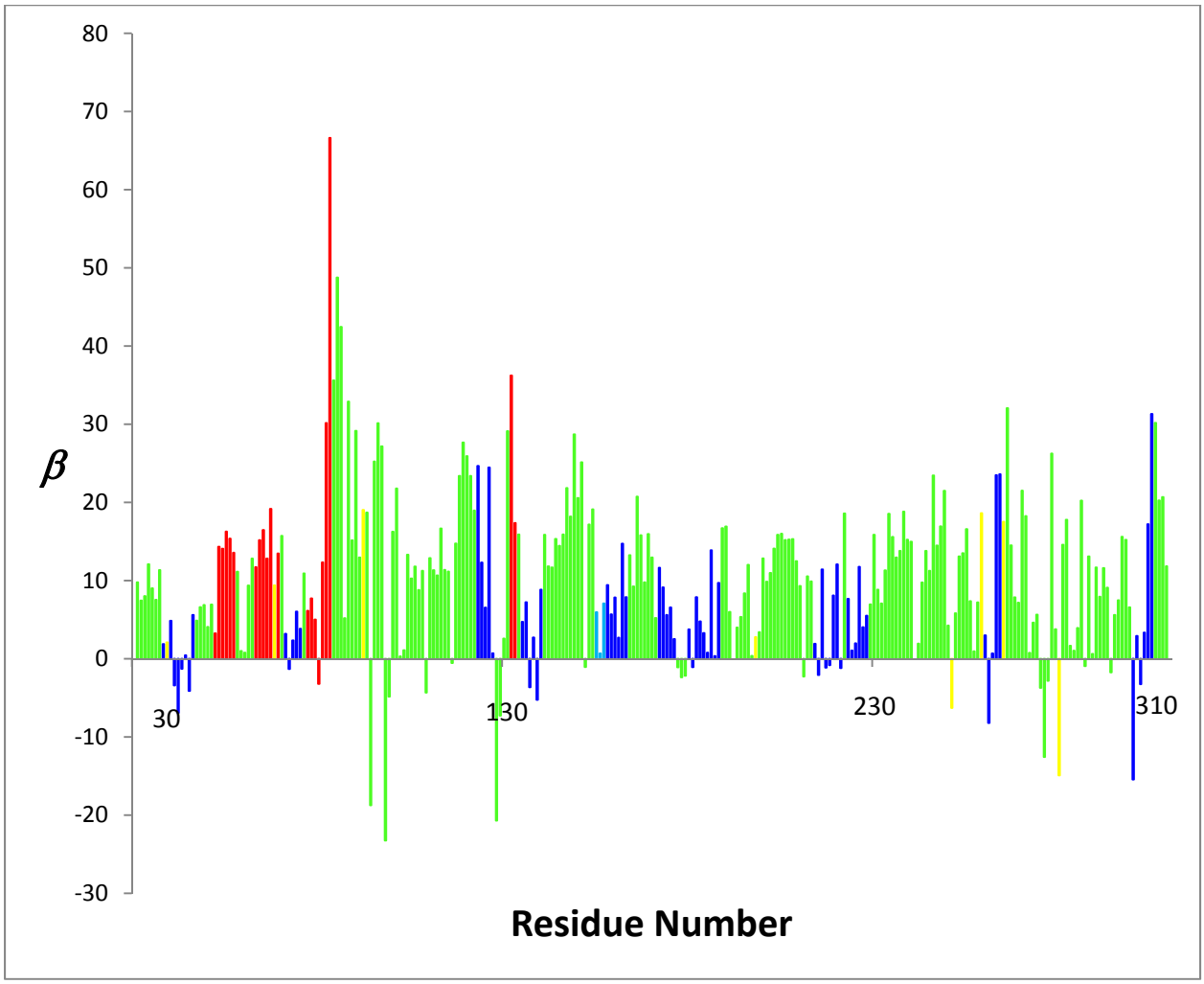


Figure 3: Chimera plot of the native (goldenrod) and first extended state (black) for the full protein. Calcium(II) ions in red, green and blue.

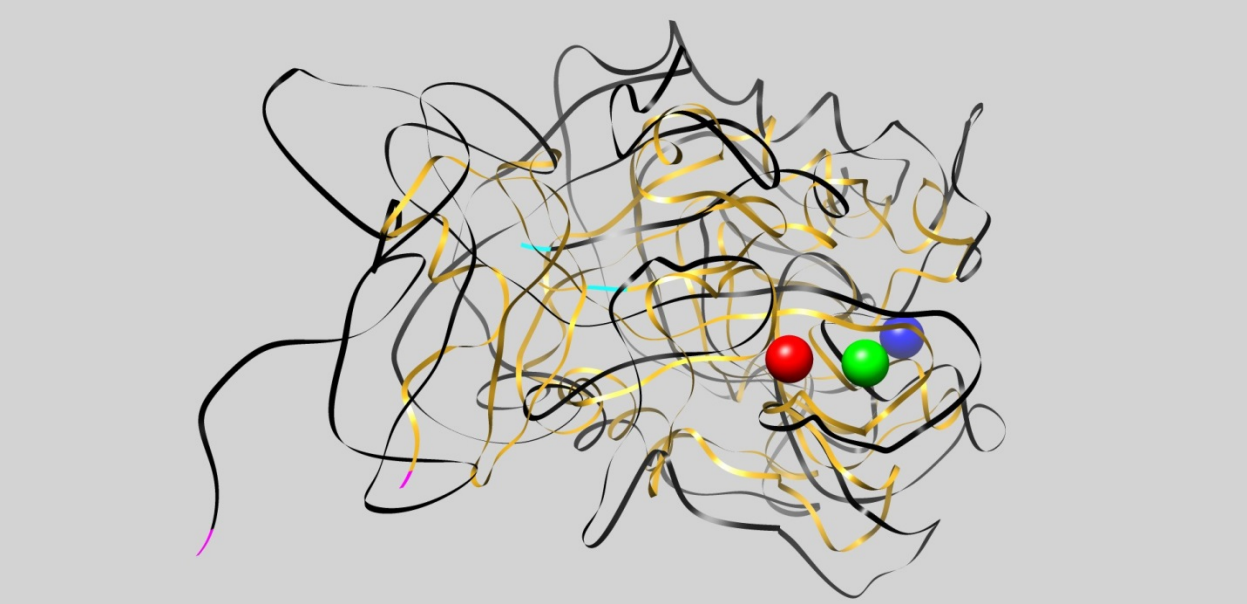


Figure 4a: Signature f vs PDB residue number for the first extended state for the (243-297) residues. The orange line represents the native state, and the magenta line is the value of <f> for the region displayed. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

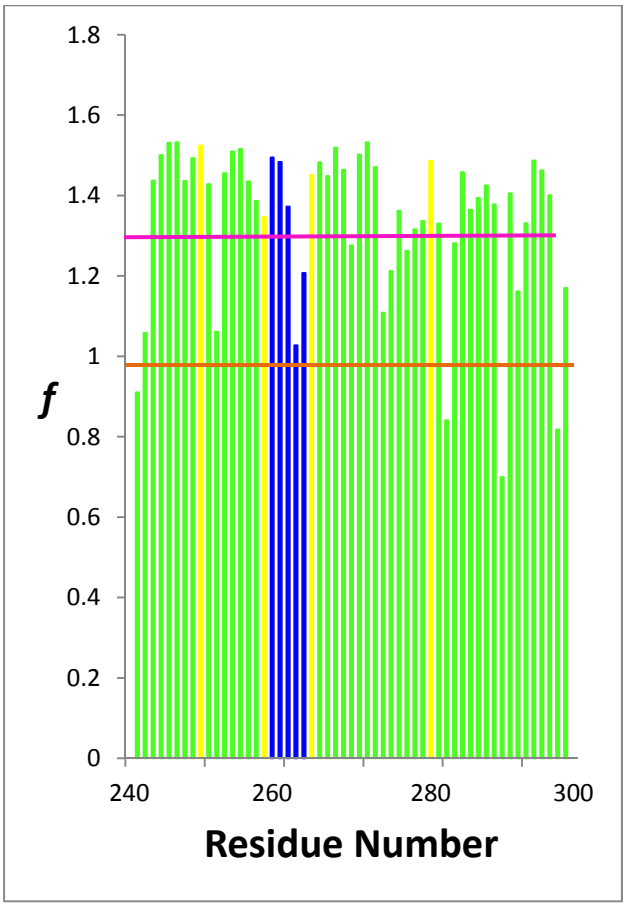


Figure 4b: Signature β vs PDB residue number for the first extended state for the (243-297) residues. Helical regions are in blue, sheets in red, SS bonding residues in yellow and all other (non-helical) regions in green.

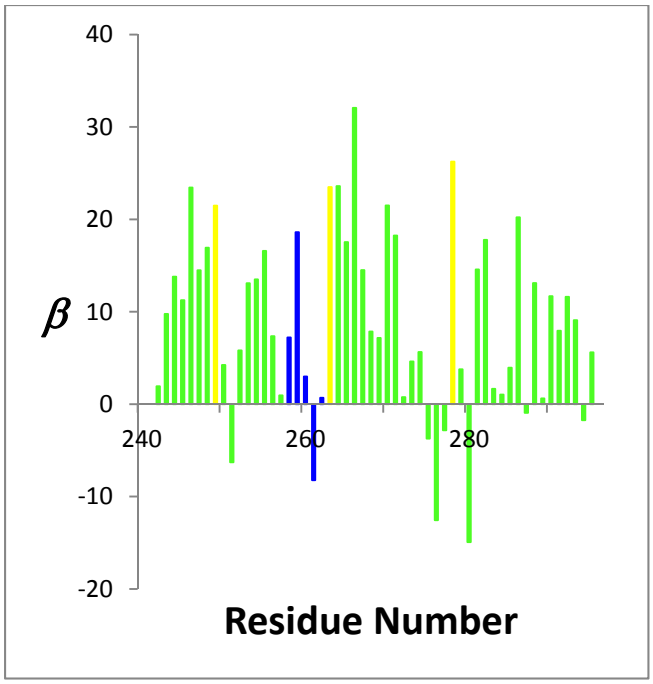
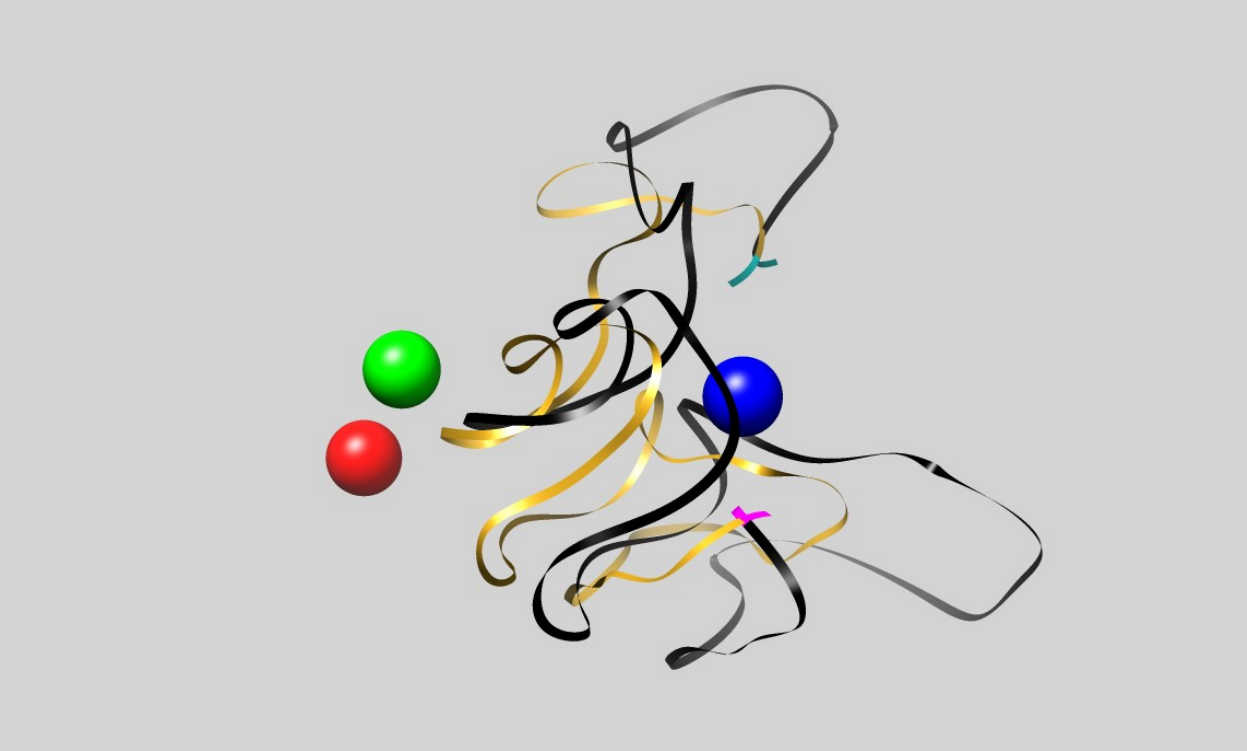


Figure 5: Chimera plot of the native (goldenrod) and first extended state (black) for the (243-297) residues. Calcium(II) ions in red, green and blue.



"Table of Contents Graphic"